



Cloudera Training for Apache HBase

Varighed: 2.00 Days **Kursus Kode: GK3904**

Beskrivelse:

Learn to deploy and manage Apache HBase in your Hadoop environment. HBase is an open-source, non-relational, distributed database that provides a fault-tolerant, scalable way to store massive quantities of data. In this course, Hadoop developers and administrators will gain the skills needed to install and maintain HBase and develop client code.

Målgruppe:

this course is suitable for Developers familiar with Apache Hadoop.

Agenda:

- HBase architecture
 - Use the HBase shell to directly manipulate HBase tables
 - Design optimal HBase schemas for efficient data storage and recovery
 - Connect to HBase using the Java API
 - Bulk-load data into HBase using MapReduce
 - Administer a HBase cluster
 - Resolve performance bottlenecks
-

Forudsætninger:

- Familiarity with Hadoop's architecture and APIs
 - Experience writing basic applications
 - Prior programming experience, preferably Java
 - Experience with databases and data modeling is helpful, but it is not required
-

Yderligere Kurser:

- There are no follow-ons for this course.
-

Indhold:

1. HBase

2. Data Model

- Tables, Row Keys, and Column Families
- Choosing Column Attributes
- Version and HBase Operations

3. HBase Shell

- Creating and Manipulating Data Using the Command-Line Shell

4. Cluster Architecture

- HMaster, RegionServers, and Zookeeper
- Compactions in HBase
- Crash Recovery

5. Storage Architecture

- Client Caching
- Data Storage and Bloom Filters
- Modifying Rows

6. Schema Design

- Creating Column Families
- Designing for Locality and Access Patterns
- Detecting and Preventing Hot Spots

7. HBase API

- Connecting to HBase Using the Java API
- Administrative Actions Using the Java API
- Accessing Data Using the Java API

8. MapReduce and Bulk Loads

- MapReduce Integration
- Bulk-Load into HBase

9. HBase Configuration

- Standalone and Distributed Run Modes
- Required Zookeeper Configurations
- Required Configuration Settings

10. HBase Administration

- Monitoring HBase Processes
- Performing HBase Backups
- Planning for HBase Capacity

11. Performance Tuning

- Preventing Network Bandwidth Bottlenecks
- Java Garbage Collection and HBase Operations
- Tuning for Client Operations
- Logging Locations and Troubleshooting Tools

Labs

- Lab 1: Use the HBase Shell
- Lab 2: Flushes and Compactions
- Lab 3: View and Read HFiles
- Lab 4: Detect Hot Spots
- Lab 5: Write an HBase Program
- Lab 6: HBase Input MapReduce
- Lab 7: Bulk Import into HBase
- Lab 8: Install Distributed HBase
- Lab 9: Pre-Split Regions

Flere Informationer:

For yderligere informationer eller booking af kursus, kontakt os på tlf.nr.: 44 88 18 00

training@globalknowledge.dk

www.globalknowledge.dk

Global Knowledge, Stamholmen 110, 2650 Hvidovre